PCT/GB 2003/005271

PATENTS • DESIGNS • COPYRIGHT • TRADE MARKS
The Patent Office

INVESTOR IN PEOPLE

GB03/05271

The Patent Office
Concept House
Cardiff Road
Newport
South Wales
NP10 8QQ

I, the undersigned, being an officer duly authorised in accordance with Section 74(1) and (4) of the Deregulation & Contracting Out Act 1994, to sign and issue certificates on behalf of the Comptroller-General, hereby certify that annexed hereto is a true copy of the documents as originally filed in connection with the patent application identified therein.

In accordance with the Patents (Companies Re-registration) Rules 1982, if a company named in this certificate and any accompanying documents has re-registered under the Companies Act 1980 with the same name as that with which it was registered immediately before re-registration save for the substitution as, or inclusion as, the last part of the name of the words "public limited company" or their equivalents in Welsh, references to the name of the company in this certificate and any accompanying documents shall be treated as references to the name with which it is so re-registered.

In accordance with the rules, the words "public limited company" may be replaced by p.l.c., plc, P.L.C. or PLC.

Re-registration under the Companies Act does not constitute a new legal entity but merely subjects the company to certain additional company law rules.

Signed

Dated 19 January 2004

PRIORITY DOCUMENT

SUBMITTED OR TRANSMITTED IN COMPLIANCE WITH RULE 17.1(a) OR (b)

Patents Form 1/77

Patents Act 1977
(Rule 16)

THE PATENT OFFICE
G
07 DEC 2002
NEWPORT

The Patent Office

09DEC02 E769518-1 C92117
P01/7700 0.00-0228614.4

Request for grant of a patent

(See the notes on the back of this form. You can also get an explanatory leaflet from the Patent Office to help you fill in this form)

The Patent Office

Cardiff Road
Newport
South Wales
NP10 8QQ

1. Your reference

Patent5

2. Patent application number
(The Patent Office will fill in this part)

0228614.4

3. Full name, address and postcode of the or of each applicant *(underline all surnames)*

Dr Guoliang Fu
24 Hayes Close, Marston
Oxford, OX3 0DZ

Patents ADP number *(if you know it)*

If the applicant is a corporate body, give the country/state of its incorporation

8422016002

4. Title of the invention

OLIGONUCLEOTIDE GUIDED ANALYSIS OF GENE EXPRESSION

5. Name of your agent *(if you have one)*

"Address for service" in the United Kingdom to which all correspondence should be sent *(including the postcode)*

Dr Guoliang Fu
24 Hayes Close, Marston
Oxford, OX3 0DZ

Patents ADP number *(if you know it)*

	Country	Priority application number *(if you know it)*	Date of filing *(day / month / year)*
6. If you are declaring priority from one or more earlier patent applications, give the country and the date of filing of the or of each of these earlier applications and *(if you know it)* the or each application number			

	Number of earlier application	Date of filing *(day / month / year)*
7. If this application is divided or otherwise derived from an earlier UK application, give the number and the filing date of the earlier application		

8. Is a statement of inventorship and of right to grant of a patent required in support of this request? *(Answer 'Yes' if:*

a) any applicant named in part 3 is not an inventor, or
b) there is an inventor who is not named as an applicant, or
c) any named applicant is a corporate body.
See note (d))

Patents Form 1/77

9. Enter the number of sheets for any of the following items you are filing with this form. Do not count copies of the same document

Continuation sheets of this form	0
Description	20
Claim *(s)*	3
Abstract	1
Drawing *(s)*	8 + 8

10. If you are also filing any of the following, state how many against each item.

Priority documents	
Translations of priority documents	
Statement of inventorship and right to grant of a patent *(Patents Form 7/77)*	
Request for preliminary examination and search *(Patents Form 9/77)*	
Request for substantive examination *(Patents Form 10/77)*	
Any other documents *(please specify)*	

11. I/~~We~~ request the grant of a patent on the basis of this application.

Signature Guoliang Fu Date 4-12-02

12. Name and daytime telephone number of person to contact in the United Kingdom

Dr Guoliang Fu
01865 724027

Warning

After an application for a patent has been filed, the Comptroller of the Patent Office will consider whether publication or communication of the invention should be prohibited or restricted under Section 22 of the Patents Act 1977. You will be informed if it is necessary to prohibit or restrict your invention in this way. Furthermore, if you live in the United Kingdom, Section 23 of the Patents Act 1977 stops you from applying for a patent abroad without first getting written permission from the Patent Office unless an application has been filed at least 6 weeks beforehand in the United Kingdom for a patent for the same invention and either no direction prohibiting publication or communication has been given, or any such direction has been revoked.

Notes

a) If you need help to fill in this form or you have any questions, please contact the Patent Office on 08459 500505.

b) Write your answers in capital letters using black ink or you may type them.

c) If there is not enough space for all the relevant details on any part of this form, please continue on a separate sheet of paper and write "see continuation sheet" in the relevant part(s). Any continuation sheet should be attached to this form.

d) If you have answered 'Yes' Patents Form 7/77 will need to be filed.

e) Once you have filled in the form you must remember to sign and date it.

f) For details of the fee and ways to pay please contact the Patent Office.

OLIGONUCLEOTIDE GUIDED ANALYSIS OF GENE EXPRESSION

Abstract

The present invention relate to methods and compositions for simultaneously analyzing multiple different polynucleotides of a polynucleotide composition comprising multiple diverse polynucleotide sequences. The subject methods and compositions may also be applied to analyze or identify single polynucleotides; however, the subject methods and compositions are particularly useful for analyzing large diverse populations of polynucleotides. Most embodiments of the invention involve hybridizing guide oligonucleotides to total RNA, genomic DNA, or cDNA for analysis, and subsequently joining parts of digested double stranded or partially double stranded guide oligonucleotides to each other. The guide oligonucleotides may be marked in identifier sequence region and constant regions so as to facilitate the simultaneous testing of multiple polynucleotides for the presence of many possible nucleotide base sequences. The identity or expression of a particular polynucleotide of interest may be ascertained by producing and quantifying a short identifier sequence derived from guide oligonucleotides after target specific hybridization. Multiple identification sequences may be obtained in parallel, thereby permitting the rapid characterization of a large number of diverse polynucleotides.

Inventor: Fu, Guoliang
24 Hayes Close
Marston
Oxford OX3 0DZ
Tel: 01865 724027

References cited [referenced By]

U.S. Patent Documents

6258539	Jul., 2001	Hunkapiller et al..
5695937	Dec., 1997	Kinzler et al..
6136537	Oct., 2000	Macevicz.

FIELD OF THE INVENTION

The invention relates generally to methods and compositions for quantitative analysis of gene expression, and more particularly, to methods and compositions for accumulating and analyzing oligonucleotide guided sequence tags sampled from a population of expressed genes.

BACKGROUND

The desire to decode the human genome and to understand the genetic basis of disease and a host of other physiological states associated differential gene expression has been a key driving force in the development of improved methods for analyzing and sequencing DNA. The human genome is estimated to contain about 30,000 genes, about 15-30% of which are active in any given tissue. Such large numbers of expressed genes make it difficult to track changes in expression patterns by available techniques, such as with hybridization of gene products to microarrays, direct sequence analysis, or the like. More commonly, expression patterns are initially analyzed by lower resolution techniques, such as differential display, indexing, subtraction hybridization, or one of the numerous DNA fingerprinting techniques, e.g. Vos et al, Nucleic Acids Research, 23: 4407-4414 (1995); Hubank et al, Nucleic Acids Research, 22: 5640-5648 (1994); Lingo et al, Science, 257: 967-971 (1992); Erlander et al, International patent application PCT/US94/13041; McClelland et al, U.S. Pat. No. 5,437,975; Unrau et al, Gene, 145: 163-169 (1994); Hubank et al, Nucleic Acids Research, 22: 5640-5648 (1994); Geng et al, BioTechniques, 25: 434-438 (1998); and the like. Higher resolution analysis is then frequently carried out on subsets of cDNA clones identified by the application of such techniques, e.g. Linskens et al, Nucleic Acids Research, 23: 3244-3251 (1995).

Recently, two techniques have been implemented that attempt to provide direct sequence information for analyzing patterns of gene expression. One involves the use of microarrays of oligonucleotides or polynucleotides for capturing complementary polynucleotides from expressed genes, e.g. Schena et al, Science, 270: 467-469 (1995); DeRisi et al, Science, 278: 680-686 (1997); Chee et al, Science, 274: 610-614 (1996); and the other involves the excision and concatenation of short sequence tags from cDNAs, followed by conventional sequencing of the concatenated tags, i.e. serial analysis of gene expression (SAGE), e.g. Velculescu et al, Science, 270: 484-486 (1995); Zhang et al, Science, 276: 1268-1272 (1997); Velculescu et al, Cell, 88: 243-251 (1997). Both techniques have shown promise as potentially robust systems for analyzing gene expression; however, there are still technical issues that need to be addressed for both approaches. For example, in microarray systems, genes to be monitored must be known and isolated beforehand, and with respect to current generation microarrays, the systems lack the complexity to provide a comprehensive analysis of mammalian gene expression, they are not readily re-usable, and they require expensive specialized data collection and analysis systems, although these of course may be used repeatedly. In SAGE systems, although no special instrumentation is necessary and an extensive installed base of DNA sequencers may be used, the selection of type IIs tag-generating enzymes is limited, and the length (nine nucleotides) of the sequence tag in current protocols severely limits the number of cDNAs that can be uniquely labeled. It can be shown that for organisms expressing large sets of genes, such as mammalian cells, the likelihood of nine-nucleotide tags being distinct for all expressed genes is extremely low. Another big issue for SAGE is that a large portion of cost and time are spent on sequencing uninformative sequence tags eg those are derived from high abundant house keeping genes.

It is clear from the above that there is a need for a technique to quickly and inexpensively analyze gene expression. The availability of such techniques would find immediate application in medical and scientific research, drug discovery, and genetic analysis in a host of applied fields.

SUMMARY OF THE INVENTION

The present invention relate to methods and compositions for simultaneously analyzing multiple different polynucleotides of a polynucleotide composition comprising multiple diverse polynucleotide sequences. The subject methods and compositions may also be applied to analyze or identify single polynucleotides; however, the subject methods and compositions are particularly useful for analyzing large diverse populations of polynucleotides. Most embodiments of the invention involve hybridizing guide oligonucleotides to total RNA, genomic DNA, or cDNA for analysis, and subsequently joining the digested double stranded or partially double stranded guide oligonucleotides to each other. The guide oligonucleotides may be marked in identifier sequence region and constant regions so as to facilitate the simultaneous testing of multiple polynucleotides for the presence of many possible nucleotide base sequences. The identity or expression of a particular polynucleotide of interest may be ascertained by producing and quantifying a short identifier sequence derived from guide oligonucleotides after target specific hybridization. Multiple identification sequences may be obtained in parallel, thereby permitting the rapid characterization of a large number of diverse polynucleotides.

Analysis of polynucleotide populations in accordance with methods of the invention may be used to provide one or more of the following types of information: (1) the nucleotide sequence of one or more polynucleotides in a complex polynucleotide composition, or (2) the relative concentrations of one or more different polynucleotides in a complex polynucleotide composition. Analysis of large complex populations of polynucleotides by the subject methods may be used to produce sufficient information about a polynucleotide population that differences between polynucleotide populations may be ascertained. Thus in some embodiments of the invention, "fingerprints" of a given polynucleotide population may be compared with "fingerprints" of other complex polynucleotide populations so as to determine differences in gene expression between the two populations. An important example of a polynucleotide composition that may be

analyzed by the invention is a cDNA preparation derived from an RNA population. The analysis of polynucleotide mixtures, particularly cDNA preparations, has numerous practical uses such as measuring gene expression for diagnostic or research purposes. Of particular interest are embodiments of the present invention that permit the majority of different polynucleotides in an RNA population may be detected.

The identity or expression of a particular polynucleotide sequence (or gene) of interest may be ascertained by producing a short identifier sequence derived by combining from the nucleotide base sequence information obtained from the hybridization of a guide oligonucleotide of known base sequence on a polynucleotide of interest.

In a typical embodiment, a guide oligonucleotide hybridizes to a target polynucleotide. An identifier sequence on the guide oligonucleotide may be isolated and determined. Multiple identifier sequences may be obtained in parallel, thereby permitting the rapid characterization of a large number of diverse polynucleotides.

BRIEF DESCRIPTION OF DRAWINGS

FIG. 1 is schematic diagram showing guide oligonucleotide. The functional regions of the guide oligonucleotide are indicated.

FIG. 2 is a schematic representation of a method of analyzing complex polynucleotides in accordance with the methods of the invention. Two or more sets pf guide oligonucleotides with different constant regions are incubated with target RNA. Target specific hybridization between guide oligos and target RNA occurs under optimal hybridization condition. Optionally, following target specific hybridization, biotinylated guide oligos are bound to avidin immobilized on a solid support and undergo stringency washing. The double stranded RNA/DNA hybrid is nicked on the RNA strand by RNase H digestion. A DNA polymerase acts on the nicked strand to catalyze 3' end extension of nicked strand. The double stranded biotinylated guide oligos are bound to avidin immobilized on a solid support and undergo stringency washing. The parts with identifier

sequence and constant region of guide oligos are released from solid support by first restriction enzyme digestion. The released parts of guide oligos can be detected directly by various methods such as mass spectrometry, electrophoresis and microarray. Alternatively, two released parts are randomly jointed together by ligation using a DNA ligase. The jointed parts are amplified by PCR or other amplification method using primers complementary to constant regions. After amplification, the amplicons are digested by first and second restriction enzymes and followed by detection with various methods for example mass spectrometry. Alternatively, the amplicons are digested by second restriction enzyme, followed by concatenation of identifier sequences by ligation, and subsequently identifier's identities are determined by sequencing or by cloning and sequencing.

FIG. 3 is a schematic diagram of analyzing complex polynucleotides using guide oligonucleotides with first restriction site and identifier region complementary to target sequence. Two or more sets pf guide oligonucleotides are incubated with target RNA or DNA. Target specific hybridization between guide oligonucleotides and target RNA or DNA occurs under optimal hybridization condition. Optionally, following target specific hybridization, biotinylated guide oligos are bound to avidin immobilized on a solid support and undergo stringency washing. The double stranded RNA/DNA or DNA/DNA hybrids are digested by first restriction enzyme at first restriction site. The 3' and 5 end parts with biotin labels are bound to avidin immobilized on a solid support and undergo stringency washing. The parts with identifier sequence and constant region of guide oligos are isolated and randomly jointed together by ligation using a DNA ligase. The jointed parts are amplified by PCR or other amplification method using primers complementary to constant regions. After amplification, the amplicons are digested by first and second restriction enzymes and followed by detection with various methods for example mass spectrometry. Alternatively, the amplicons are digested by second restriction enzyme, followed by concatenation of identifier sequences by ligation, and subsequently identifier sequence's identities are determine by DNA sequencing or by cloning and sequencing.

FIG. 4 is a schematic diagram of analyzing biotinilated cDNA. cDNA is formed by reverse transcription using a reverse transcriptase and a biotinylated poly dT primer. The cDNA is split into two portions and each hybridizes to a set of guide oligonucleotide. The two sets of guide oligonucleotides have different constant regions in different orientations. The cDNA is immobilized on a solid phase by binding to avidin. The immobilized cDNA is then digested with a first restriction endonuclease and undergo stringency washing. The parts with identifier sequence and constant region of guide oligos are isolated and randomly jointed together by ligation using a DNA ligase. The jointed parts are amplified by PCR or other amplification method using primers complementary to constant regions. After amplification, the amplicons are digested by first and second restriction enzymes and followed by detection with various methods for example mass spectrometry. Alternatively, the amplicons are digested by second restriction enzyme, followed by concatenation of identifier sequences by ligation, and subsequently identifier sequence's identities are determined by DNA sequencing or by cloning and sequencing.

DESCRIPTION OF THE PREFERRED EMBODIMENTS

Guide oligonucleotide is a linear single-stranded nucleic acid molecule, generally containing between 30 to 1000 nucleotides, preferably between about 40 to 300 nucleotides, and most preferably between about 50 to 150 nucleotides. Regions of guide oligonucleotides have specific functions making the guide oligonucleotide useful for embodiments of invention. These regions are referred to as the target complementary region, constant region, identifier sequence, at least one restriction site - usually there are two restriction sites termed first and second restriction sites, with or without 5' or 3 end label. A guide oligonucleotide may not contain all regions.

1. Target complementary region

There is generally one target complementary region on guide oligonucleotide, but more than one target complementary regions will also work. The target complementary region

can be any length that supports specific and stable hybridization between the target complementary region and the target sequence. For this purpose, a length of 9 to 60 nucleotides for target complementary region is preferred, with target complementary regions 15 to 40 nucleotides long being most preferred.

If the target sequence is RNA, upon hybridization to the target complementary region, the target RNA sequence in the hybrid RNA/DNA is nicked by RNase H digestion at various non-specific sites. In one embodiment, part of the target complementary region can be made by RNA. Upon hybridization between target RNA sequence and the target complementary region, the RNA/RNA hybrid will be resistant to digestion with RNase H. This is beneficial that the target RNA in the hybrid formed between target and guide oligonucleotide will not be digested away so that nicking and extension can occur.

2. Constant regions

The constant region serves as priming site for amplification. In other word, the constant region is complementary to primer used for amplification. For this purpose, a length of 15 to 50 nucleotides for the constant region is preferred, and 18 to 35 nucleotides long are most preferred. The "constant region" is said to be constant because the constant regions of a set of guide oligonucleotides are functionally the same as each other with respect to their hybridization specificity to amplification primers as used in the methods of the invention. The constant region can have any desired sequence. In general, the sequence of the constant region can be chosen such that it is not significantly similar to any sequence in target polynucleotide.

The "constant region" of a guide oligonucleotide may be located either 5' or 3' with respect to the target complementary region. The selection of the relative orientation of the constant region with respect to the target complementary region in a given embodiment of the invention will vary in accordance with choice of which strand of target polynucleotide is selected for analysis. In some embodiments of invention, a set or several sets of guide oligonucleotides have the same constant region orientation, but the

sequences of constant regions are different between of sets of oligos (Fig. 2). In other embodiments of invention, a set or several sets of guide oligonucleotides have the different constant region orientations, as well as different sequences of constant regions between of guide oligonucleotide sets (FIG. 3 and FIG. 4).

The term "a set of guide oligonucleotides" as used herein refers to a plurality of different guide oligonucleotides used in conjunction with each other, wherein each guide oligonucleotide in the set has a functionally identical constant region, e.g., all of the constant regions are identical or have essentially the same sequence-specific hybridization properties, and the target complementary regions and the identifier sequences are different from one another.

3. Identifier sequence

Identifier sequence is located between first and second restriction sites. Identifier sequence can comprise any sequence of any length that is unique to a guide oligonucleotide. The identifier sequence serves as a role to distinguish individual guide oligonucleotides. For this purpose, a length of 4 to 30 nucleotides for the identifier sequence is preferred, and 5 to 20 nucleotides long are most preferred. The identifier sequence can have any desired sequence. In some embodiments of the invention, the identifier sequence and first restriction site is a contiguous to target complementary region. In other words, the target complementary region, first restriction enzyme and identifier sequence act as a whole to hybridize to a target sequence. In other embodiments of the invention, the identifier sequence can be randomly chosen, and may not contain any significant similar sequence in target polynucleotide. All identifier sequences of the guide oligonucleotides in a set are not needed to be the same length. Actually, the identity of a identifier sequence is determined by both its length and sequence.

A identifier sequence may be specifically associated with a given oligonucleotide, the base sequence of the nucleotide may be determined because of the predetermined correlation between the base sequence of the oligonucleotide and the identifier sequence.

4. First and second restriction enzyme sites

Any restriction enzyme sites can be used as first and second restriction enzyme sites. In general, four base cutters are preferred, and first and second restriction sites are not same. In some embodiments of the invention, the identifier sequence and first restriction site is contiguous to target complementary region. In other words, the target complementary region, first restriction enzyme and identifier sequence act as a whole to hybridize a target sequence.

5. 5' or 3' end label

Extension of a guide oligonucleotide by a polymerase may be blocked by a blocking group at its 3' end. The blockage of 3' end of guide oligonucleotide can be achieved by any means known in the art. Blocking groups are chemical moieties which can be added to a nucleic acid to inhibit nucleic acid polymerization catalyzed by a nucleic acid polymerase. Blocking groups are typically located at the terminal 3' end of guide oligonucleotide which is made up of nucleotides or derivatives thereof. By attaching a blocking group to a terminal 3' OH, the 3' OH group is no longer available to accept a nucleoside triphosphate in a polymerization reaction.

Numerous different groups can be added to block the 3' end of a probe sequence. Examples of such groups include alkyl groups, non-nucleotide linkers, phosphorothioate, alkane-diol residues, peptide nucleic acid, and nucleotide derivatives lacking a 3' OH (e.g., cordycepin).

An alkyl blocking group is a saturated hydrocarbon up to 12 carbons in length which can be a straight chain or branched, and/or contain a cyclic group. More preferably, the alkyl blocking group is a C.sub.2 -C.sub.6 alkyl which can be a straight chain or branched, and/or contain a cyclic group.

In certain embodiments, guide oligonucleotide can include one or more moieties incorporated into 5' or 3' terminus or internally of guide oligonucleotide that allow for the affinity separation of products derived from guide oligonucleotide associated with the label from unassociated parts. Preferred capture moieties are those that can interact specifically with a cognate ligand. For example, capture moiety can include biotin, digoxigenin etc. Other examples of capture groups include ligands, receptors, antibodies, haptens, enzymes, chemical groups recognizable by antibodies or aptamers. The capture moieties can be immobilized on any desired substrate. Examples of desired substrates include, e.g., particles, beads, magnetic beads, optically trapped beads, microtiter plates, glass slides, papers, test strips, gels, other matrices, nitrocellulose, nylon. For example, when the capture moiety is biotin, the substrate can include streptavidin.

In many embodiments of the invention, multiple guide oligonucleotides are selected to be used in conjunction with one another, i.e., sets of guide oligonucleotides, thereby providing for the simultaneous analysis of multiple polynucleotides when the different oligonucleotides are used in conjunction with one another.

The term "oligonucleotide" as used herein is used broadly to refer to any naturally occurring nucleic acid, or any synthetic analogs thereof, that have the chemical properties required for use in the subject methods, e.g., the ability to sequence specifically hybridize different polynucleotides. Thus, examples of oligonucleotides include DNA, RNA, phosphorthioates PNAs (peptide nucleic acids), phosphoramidates and the like. Method for synthesizing oligonucleotides are well known to those skilled in the art, examples of such synthesis can be found for example in U.S. Pat. Nos. 4,419,732; 4,458,066; 4,500,707; 4,668,777; 4,973,679; 5,278,302; 5,153,319; 5,786,461; 5,773,571; 5,539,082; 5,476,925; and 5,646,260.

The term "joining" as used herein, with respect to oligonucleotides or polynucleotides refers to the covalent attachment of two separate nucleic acids to produce a single larger nucleic acid with a contiguous backbone. Preferred methods of polynucleotide joining are ligase (e.g., T4 ligase) catalyzed reactions. However, non-enzymatic ligation methods

may also be employed. Examples of ligation reactions that are non-enzymatic include the non-enzymatic ligation techniques described in U.S. Pat. Nos. 5,780,613 and 5,476,930, which are herein incorporated by reference.

The term "fingerprint" as used herein refers to a set of data relating to a complex polynucleotide population in which the relative concentrations of the different polynucleotide that formed the population are measured.

Enzymes

For some embodiments of invention, a DNA polymerase is used to extend nicked strand. Any DNA polymerase can be used, preferred DNA polymerases include E.coli DNA polymerase I and its derivatives, Taq DNA polymerase, T7 DNA polymerase, Bst polymerase.

The disclosed method makes use of restriction enzymes (also referred to as restriction endonucleases) for cleaving double stranded nucleic acids. Other nucleic acid cleaving reagents also can be used. Preferred nucleic acid cleaving reagents are those that cleave nucleic acid molecules in a sequence-specific manner. Many restriction enzymes are known and can be used with the disclosed method. Restriction enzymes generally have a recognition sequence and a cleavage site. Restriction enzymes are widely available commercially, and procedures for using them are well known to persons of ordinary skill in the art of molecular biology. Suitable restriction endonucleases may produce either blunt ends or overhanging ends.

Type IIs restriction endonucleases may also be used as a restriction endonuclease in oligonucleotide guided restriction endonuclease digestion. Type IIs restriction endonuclease have recognition sites that are different than the cleavage site. Type IIs restriction endonucleases are of particular interest because they may be used to produce small restriction fragments of a uniform size because the property of type IIs enzymes to cleave at a fixed distance from the recognition site, irrespective of the cleavage sequence.

To use a type IIs restriction endonuclease, a second restriction site having type IIs restriction endonuclease recognition site may be employed.

In one embodiment, nicking the hybridized target RNA at predetermined RNA sequences is carried out with a double-stranded ribonuclease. Such ribonucleases nick or excise ribonucleic acid sequences from double-stranded RNA/DNA hybridized strands. An example of a ribonuclease useful in the practice of this invention is RNase H. RNase H is a RNA specific digestion enzyme which cleaves RNA found in DNA/RNA hybrids in a non-sequence-specific manner. Other ribonucleases and enzymes may be suitable to nick or excise RNA from RNA/DNA strands, such as Exo III and reverse transcriptase.

In one embodiment, a single stranded cDNA is used as target source (FIG. 4). cDNA is formed by reverse transcription using a reverse transcriptase and a biotinylated poly dT primer. The cDNA is immobilized on a solid phase by binding to avidin. The cDNA may be split into two portions and each hybridize to a set of guide oligonucleotide. After target specific hybridization with guide oligonucleotide, the immobilized cDNA is then digested with a first restriction endonuclease. The released, i.e., not bound, restriction fragments may then be isolated or washed away.

The materials described above can be packaged together in any suitable combination as a kit useful for performing the disclosed method.

The present invention relate to methods and compositions for simultaneously analyzing multiple different polynucleotides of a polynucleotide composition comprising multiple diverse polynucleotide sequences. The subject methods and compositions may also be applied to analyze or identify single polynucleotides; however, the subject methods and compositions are particularly useful for analyzing large diverse populations of polynucleotides. Most embodiments of the invention involve hybridizing guide oligonucleotides to total RNA, genomic DNA, or cDNA for analysis, and subsequently joining the digested double stranded or partially double stranded guide oligonucleotides to each other. The guide oligonucleotides may be marked in identifier sequence region

and constant regions so as to facilitate the simultaneous testing of multiple polynucleotides for the presence of many possible nucleotide base sequences. The identity or expression of a particular polynucleotide of interest may be ascertained by producing and quantifying a short identifier sequence derived from guide oligonucleotides after target specific hybridization. Multiple identification sequences may be obtained in parallel, thereby permitting the rapid characterization of a large number of diverse polynucleotides.

Analysis of polynucleotide populations in accordance with methods of the invention may be used to provide one or more of the following types of information: (1) the nucleotide sequence of one or more polynucleotides in a complex polynucleotide composition, or (2) the relative concentrations of one or more different polynucleotides in a complex polynucleotide composition. Analysis of large complex populations of polynucleotides by the subject methods may be used to produce sufficient information about a polynucleotide population that differences between polynucleotide populations may be ascertained. Thus in some embodiments of the invention, "fingerprints" of a given polynucleotide population may be compared with "fingerprints" of other complex polynucleotide populations so as to determine differences in gene expression between the two populations. An important example of a polynucleotide composition that may be analyzed by the invention is a cDNA preparation derived from an RNA population. The analysis of polynucleotide mixtures, particularly cDNA preparations, has numerous practical uses such as measuring gene expression for diagnostic or research purposes. Of particular interest are embodiments of the present invention that permit the majority of different polynucleotides in an RNA population may be detected.

The identity or expression of a particular polynucleotide sequence (or gene) of interest may be ascertained by producing a short identifier sequence derived by combining from the nucleotide base sequence information obtained from the hybridization of a guide oligonucleotide of known base sequence on a polynucleotide of interest.

In a typical embodiment, a guide oligonucleotide hybridizes to a target polynucleotide. An identifier sequence on the guide oligonucleotide may be isolated and determined. Multiple identifier sequences may be obtained in parallel, thereby permitting the rapid characterization of a large number of diverse polynucleotides.

In one embodiment, two or more sets of guide oligonucleotides with different constant regions are incubated with target RNA (FIG.2). Target specific hybridization between target complementary region of guide oligonucleotides and target RNA occurs under optimal hybridization condition. Optionally, following target specific hybridization, biotinylated guide oligos are bound to avidin immobilized on a solid support and undergo stringency washing. The double stranded RNA/DNA hybrid is nicked on the RNA strand by RNase H digestion. A DNA polymerase acts on the nicked strand to catalyze 3' end extension of nicked strand. The double stranded biotinylated guide oligos are bound to avidin immobilized on a solid support and undergo stringency washing. The parts with identifier sequence and constant region of guide oligos are released from solid support by first restriction enzyme digestion. The released parts of guide oligos can be detected directly by various methods such as mass spectrometry, electrophoresis and microarray. Alternatively, the released parts are randomly jointed together by ligation using a DNA ligase. The jointed parts are amplified by PCR using primers complementary to constant regions. After amplification, the amplicons are digested by first and second restriction enzymes and followed by detection with various methods for example mass spectrometry. Alternatively, the amplicons are digested by second restriction enzyme, followed by concatenation of identifier sequences by ligation, and subsequently identifier sequence's identities are determined by DNA sequencing or by cloning and sequencing.

In another embodiment (FIG. 3), a method is provided for analyzing complex polynucleotides using guide oligonucleotides with first restriction site and identifier region complementary to target sequence. Two or more sets pf guide oligonucleotides are incubated with target RNA or DNA. Target specific hybridization between guide oligonucleotides and target RNA or DNA occurs under optimal hybridization condition. Optionally, following target specific hybridization, biotinylated guide oligos are bound to

avidin immobilized on a solid support and undergo stringency washing. The double stranded RNA/DNA or DNA/DNA hybrids are digested by first restriction enzyme at first restriction site. The 3' and 5 end parts with biotin labels are bound to avidin immobilized on a solid support and undergo stringency washing. The parts with identifier sequence and constant region of guide oligos are isolated and randomly jointed together by ligation using a DNA ligase. The jointed parts are amplified by PCR or other amplification method using primers complementary to constant regions. After amplification, the amplicons are digested by first and second restriction enzymes and followed by detection with various methods for example mass spectrometry. Alternatively, the amplicons are digested by second restriction enzyme, followed by concatenation of identifier sequences by ligation, and subsequently identifier sequence's identities are determine by DNA sequencing or by cloning and sequencing.

In yet another embodiment (FIG. 4), a method is provided for analyzing biotinilated cDNA. cDNA is formed by reverse transcription using a reverse transcriptase and a biotinylated poly dT primer. The cDNA is split into two portions and each hybridizes to a set of guide oligonucleotide. The two sets of guide oligonucleotides have different constant regions in different orientations. The cDNA is immobilized on a solid phase by binding to avidin. The immobilized cDNA is then digested with a first restriction endonuclease and undergo stringency washing. The parts with identifier sequence and constant region of guide oligos are isolated and randomly jointed together by ligation using a DNA ligase. The jointed parts are amplified by PCR or other amplification method using primers complementary to constant regions. After amplification, the amplicons are digested by first and second restriction enzymes and followed by detection with various methods for example mass spectrometry. Alternatively, the amplicons are digested by second restriction enzyme, followed by concatenation of identifier sequences by ligation, and subsequently identifier sequence's identities are determined by DNA sequencing or by cloning and sequencing.

A. Target specific hybridization

A guide oligonucleotide or a set of guide oligonucleotides or more than one sets of guide oligonucleotides are incubated with a sample containing DNA, RNA, or both, under suitable hybridization conditions, so that a double stranded DNA/DNA or RNA/DNA in the target complementary regions of guide oligonucleotides are formed. A stringent hybridization condition allows subsequent isolation and amplification to be dependent on the perfect match between a target sequence and guide oligonucleotides so that allele discrimination can be achieved.

B. Capturing on solid support

The 3' or 5' end of guide oligonucleotide may be labeled by a capture moiety, for example biotin (FIG, 2 and FIG. 3). Alternatively, the target polynucleotide may be labeled by a capture moiety, for example, a cDNA from mRNA is formed using a biotinylated poly dT primer (FIG. 4). Upon target specific hybridization, the biotin labeled oligonucleotide or polynecleotide are bound to streptavidin on a solid support, for example a beads. A stringency washing may be carried out to remove any unspecific hybridized oligonucleotide or polynucleotide.

C. Forming functional first restriction site

If the first restriction enzyme recognition site and identifier sequence are complementary to target sequence and contiguous to target complementary region of guide oligonucleotide, a double stranded functional first restriction site is already formed after target specific hybridization between target molecule and guide oligonucleotide (FIG. 3 and FIG. 4).

In the case of that target molecule is RNA and the first restriction site is not complementary to target, a functional double stranded restriction site is created by nicking target RNA strand by RNase H digestion and extension of 3' end of nicked strand on guide oligonucleotide template by a DNA polymerase (FIG. 2). RNase H is a RNA specific digestion enzyme which cleaves RNA found in DNA/RNA hybrids in a non-

sequence-specific manner. To prevent complete digestion away of RNA strand, a portion of target complementary region of the guide oligonucleotide may be made by RNA, thus RNA/RNA hybrid is resistant to be digested by RNase H.

Optionally, if the labeled oligonucleotide or polynucleotide are not captured on a solid support after target specific hybridization, the labeled oligonucleotide or polynucleotide can be captured after RNase H digestion and polymerase extension (FIG. 2).

C. Digestion by the first restriction enzyme and isolation of digested parts

Once double stranded functional restriction site is formed, its cognate restriction enzyme acts on and cleave the double stranded nucleic acid.

Digested parts with constant region and identifier sequence are isolated by capturing appropriate parts on a solid support.

Optionally, the released parts of guide oligonucleotides with constant region and identifier sequence can be detected directly by various methods such as mass spectrometry, electrophoresis and microarray.

D. Joining digested parts with constant region and identifier sequence, amplification of jointed product

In a preferred embodiment of the invention, the digested parts with constant region and identifier sequence are joined together by DNA ligation. The amplification of the jointed fragments may be achieved through the use of primers that can anneal to constant regions of guide oligonucleotides. The product of the amplification product is referred to herein as "amplicon".

A variety of primer-dependent polynucleotide amplification techniques may be used for amplification. Such techniques include strand displacement amplification, 3SR

amplification, and the like. The polymerase chain reaction (PCR) is particularly preferred for amplifying the jointed parts. The polymerase chain reaction is described in, among other places, Diffenbach and Dveksler, PCR Primer Cold Spring Harbor Press, Cold Spring Harbor, N.Y. (1995) and U.S. Pat. Nos. 4,683,202; 4,683,195; 4,800,159; 4,965,188; and 5,333,675.

In embodiments of the invention employing polynucleotide amplification, the primers for use in the polynucleotide amplification primers are selected so as to work in conjunction with the guide oligonucleotides used in the given embodiment.

E. Detection of amplicon

The amplicon can be detected by any method known in the art. One example is that the amplicon is digested by first and second restriction enzymes and identifier sequences are detected by mass spectrometry. A preferred detection method is that the amplicons are digested by second restriction enzyme, followed by concatenation of identifier sequences by DNA ligation, and subsequently determine identifier sequence's identity by cloning and DNA sequencing.

Kits

The invention also includes kits for performing one or more of the different methods for analyzing polynucleotide population described herein. Kits generally contain two or more reagents necessary to perform the subject methods. The reagents may be supplied in pre-measured amount for individual assays so as to increase reproducibility.

In one embodiment, the subject kits comprise guide oligonucleotides and primers for use to amplify guided fragments. The kits of the invention may also include one or more additional reagents required for various embodiments of the subject methods. Such additional reagents include, but are not limited to: restriction enzymes, DNA polymerases, buffers, nucleotides, and the like.

Incorporation By Reference

All publications, patent applications, and patents referenced in the specification are herein incorporated by reference to the same extent as if each individual publication or patent application was specifically and individually indicated to be incorporated by reference.

Equivalents

All publications, patent applications, and patents mentioned in this specification are indicative of the level of skill of those skilled in the art to which this invention pertains. Although only a few embodiments have been described in detail above, those having ordinary skill in the molecular biology art will clearly understand that many modifications are possible in the preferred embodiment without departing from the teachings thereof. All such modifications are intended to be encompassed within the following claims. The foregoing written specification is considered to be sufficient to enable skilled in the art to which this invention pertains to practice the invention. Indeed, various modifications of the above-described modes for carrying out the invention which are apparent to those skilled in the field of molecular biology or related fields are intended to be within the scope of the following claims

ABSTRACT

OLIGONUCLEOTIDE GUIDED ANALYSIS OF GENE EXPRESSION

The present invention relate to methods and compositions for simultaneously analyzing multiple different polynucleotides of a polynucleotide composition comprising multiple diverse polynucleotide sequences. The subject methods and compositions may also be applied to analyze or identify single polynucleotides; however, the subject methods and compositions are particularly useful for analyzing large diverse populations of polynucleotides. Most embodiments of the invention involve hybridizing guide oligonucleotides to total RNA, genomic DNA, or cDNA for analysis, and subsequently joining parts of digested double stranded or partially double stranded guide oligonucleotides to each other. The guide oligonucleotides may be marked in identifier sequence region and constant regions so as to facilitate the simultaneous testing of multiple polynucleotides for the presence of many possible nucleotide base sequences. The identity or expression of a particular polynucleotide of interest may be ascertained by producing and quantifying a short identifier sequence derived from guide oligonucleotides after target specific hybridization. Multiple identification sequences may be obtained in parallel, thereby permitting the rapid characterization of a large number of diverse polynucleotides.

Claims

What is claimed is:

1. A method of analyzing a polynucleotide, said method comprising
(a) hybridizing guide oligonucleotides to target polynucleotide, wherein the guide oligonucleotide has target complementary region, constant region, identifier sequence, at least one restriction site, with or without 5' or 3 end label;
(b) forming double stranded or partial double stranded guide oligonucleotides;
(c) digesting said double stranded or partial double stranded guide oligonucleotides with first restriction enzyme;
(d) isolating the digested parts of said double stranded or partial double stranded guide oligonucleotide;
(e) analyzing said digested parts.

2. The method of claim 1, wherein said step of forming double stranded or partial double stranded guide oligonucleotides includes nicking a RNA strand of RNA/DNA hybrid by a nuclease, extending the nicked strand on guide oligonucleotide template by a DNA polymerase.

3. The method of claim 2, wherein said nuclease is RNase H.

4. The method of claim 1, wherein said step of analyzing the digested parts of said double stranded or partial double stranded guide oligonucleotides comprising detecting the digested parts of said double stranded or partial double stranded guide oligonucleotides by mass spectrometry, electrophoresis and microarray.

5. The method of claim 1, wherein said step of analyzing the digested parts of said double stranded or partial double stranded guide oligonucleotides comprising:
(a) ligating said digested parts;

(b) amplifying ligated products using primers that are complementary to constant regions of guide oligonucleotides;

(c) analyzing the amplified products.

6. The method of claim 5, where in said step of analyzing the amplified products comprising determining the nucleotide sequence of said amplified products.

7. The method of claim 5, where in said step of analyzing the amplified products comprising:

(a) digesting the said amplified products with first and second restriction enzymes to release individual identifier sequences;

(b) detecting and quantifying said identifier sequences by a detection method.

8. The method of claim 5, where in said step of analyzing the amplified products comprising:

(a) digesting the said amplified products with second restriction enzymes to release linked identifier sequences;

(b) ligating said linked identifier sequences to produce a concatemer;

(c) determining the nucleotide sequence of identifier sequences in said concatemer.

9. The method of claim 8, where in said step of determining the nucleotide sequence of identifier sequences in said concatemer includes cloning and sequencing.

10. The method according to claim 1 wherein the polynucleotide is a RNA, cDNA or genomic DNA.

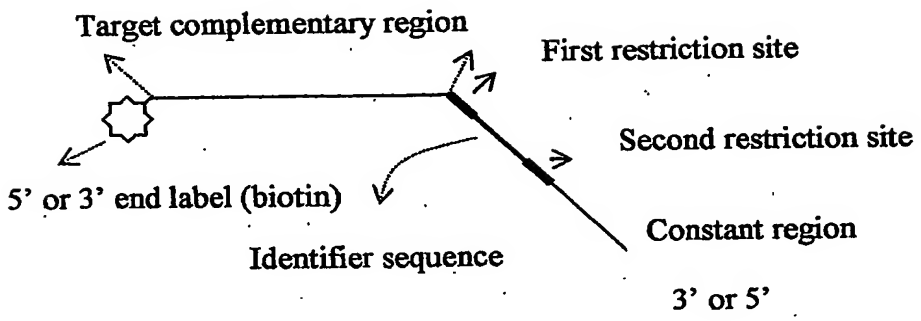
11. The method of claim 1, wherein the identifier sequence comprises any sequence of any length that is unique to a guide oligonucleotide.

12. The method of claim 1, where in said step of isolating the digested parts of said double stranded or partial double stranded guide oligonucleotide comprising

(a) immobilizing labeled polynucleotide or labeled oligonucleotide on a solid support,

(b) purifying the parts with identifier sequence and constant regions.

13. The method of claim 1, wherein the said at least one restriction site comprise first and second restriction site.



Guide oligonucleotide

FIG. 1

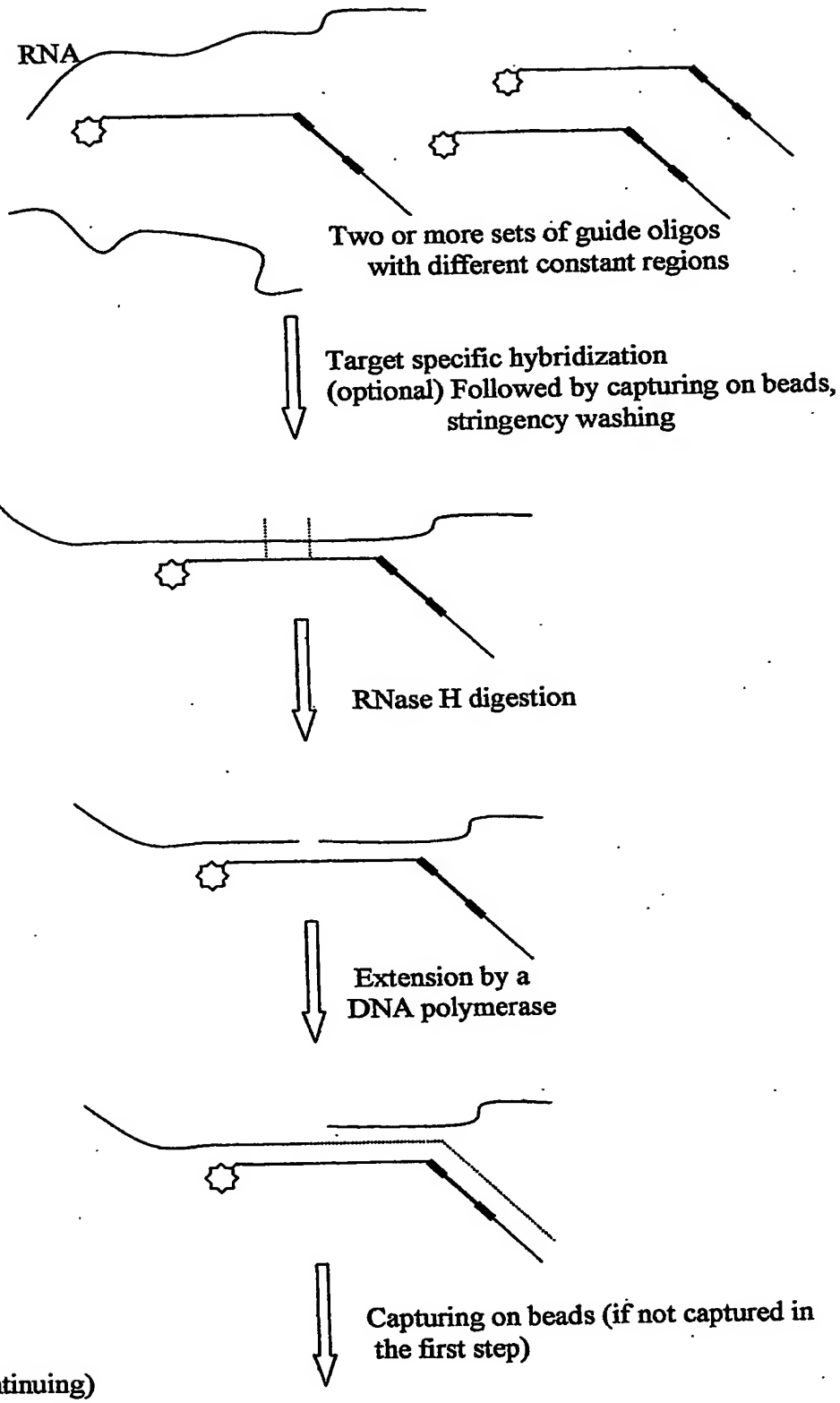


FIG.2 (continuing)

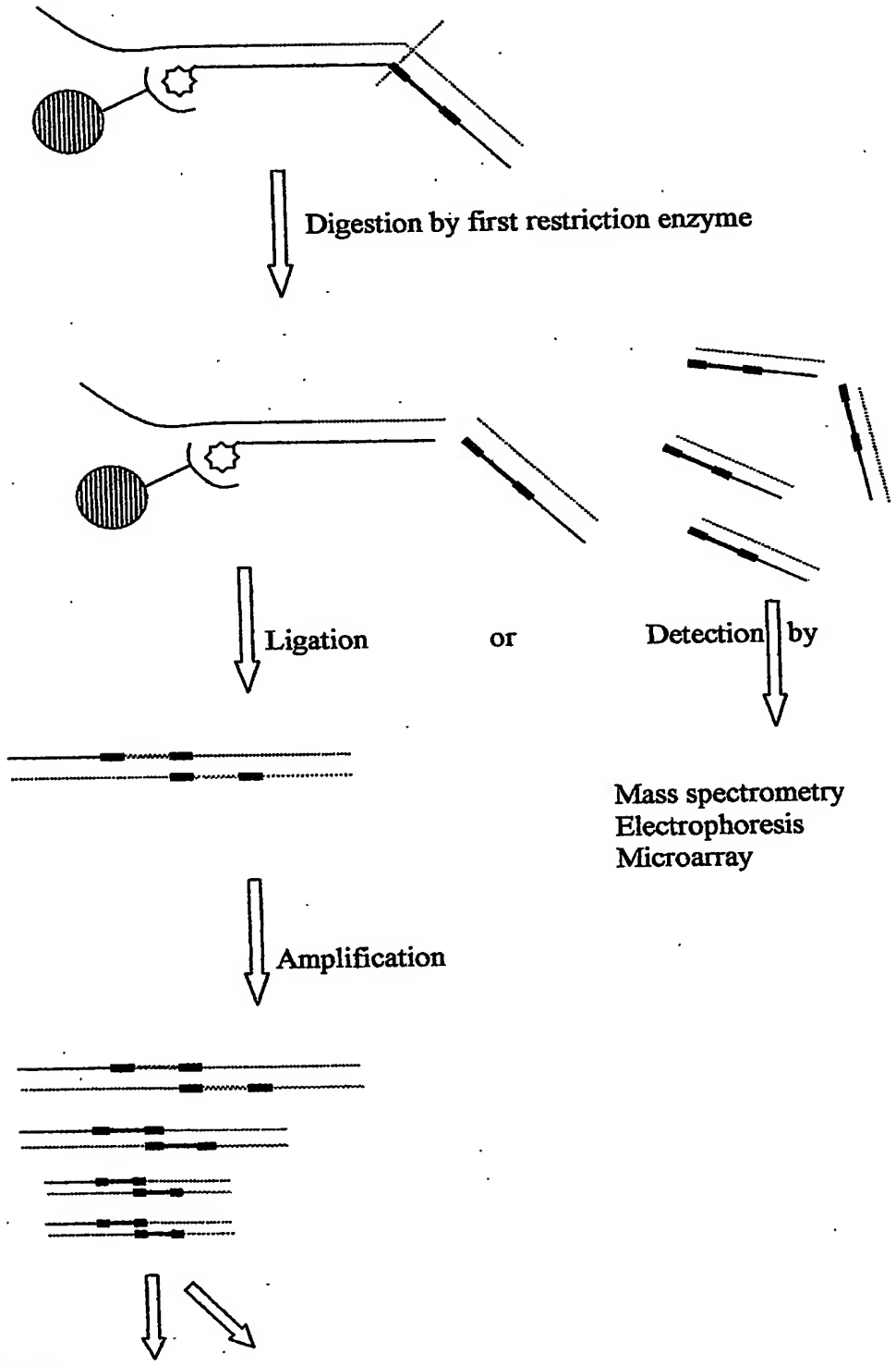


FIG. 2 (continuing)

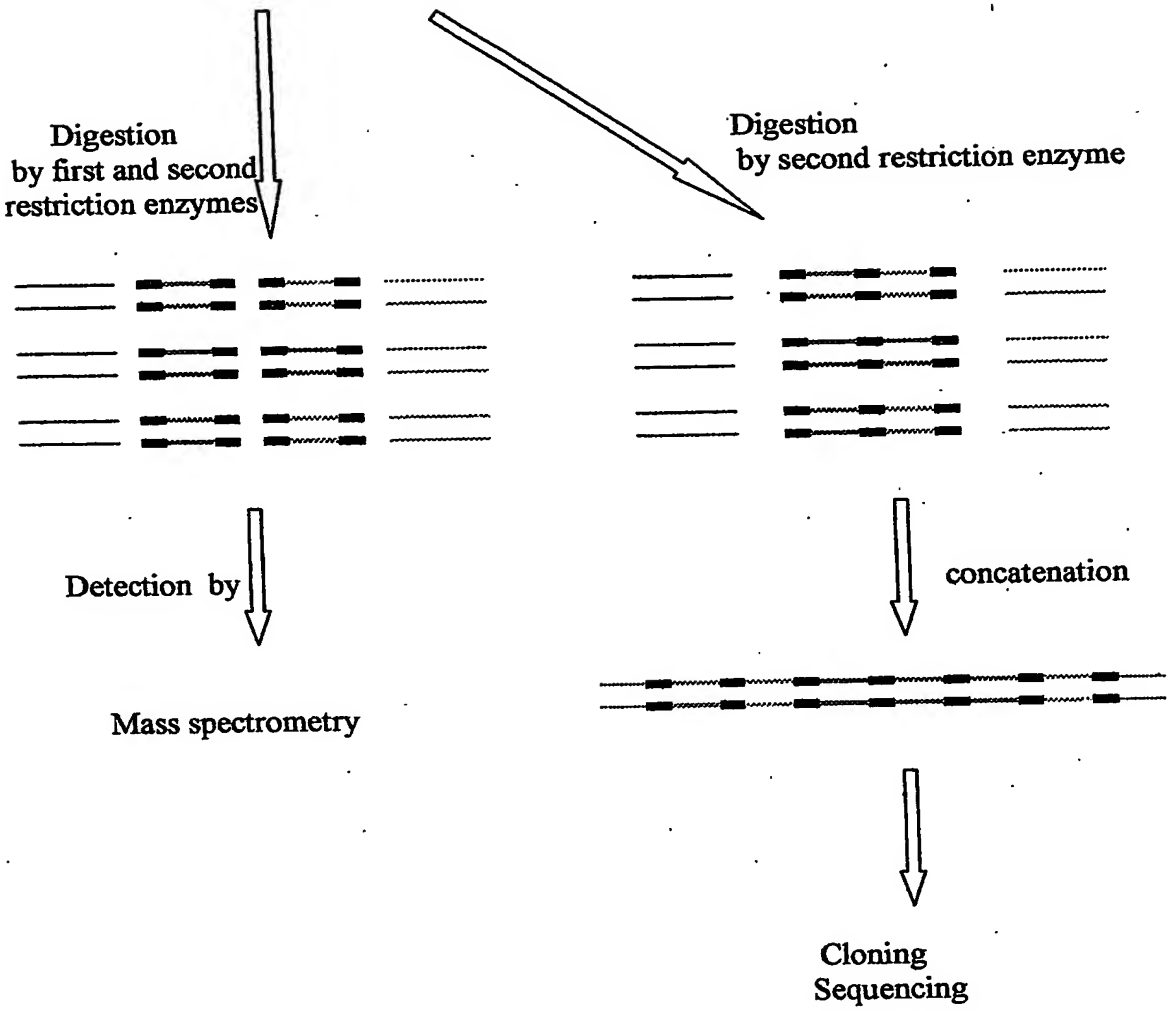

Digestion
by first and second
restriction enzymes
Digestion
by second restriction enzyme
Detection by
Mass spectrometry
concatenation
Cloning
Sequencing

FIG. 2

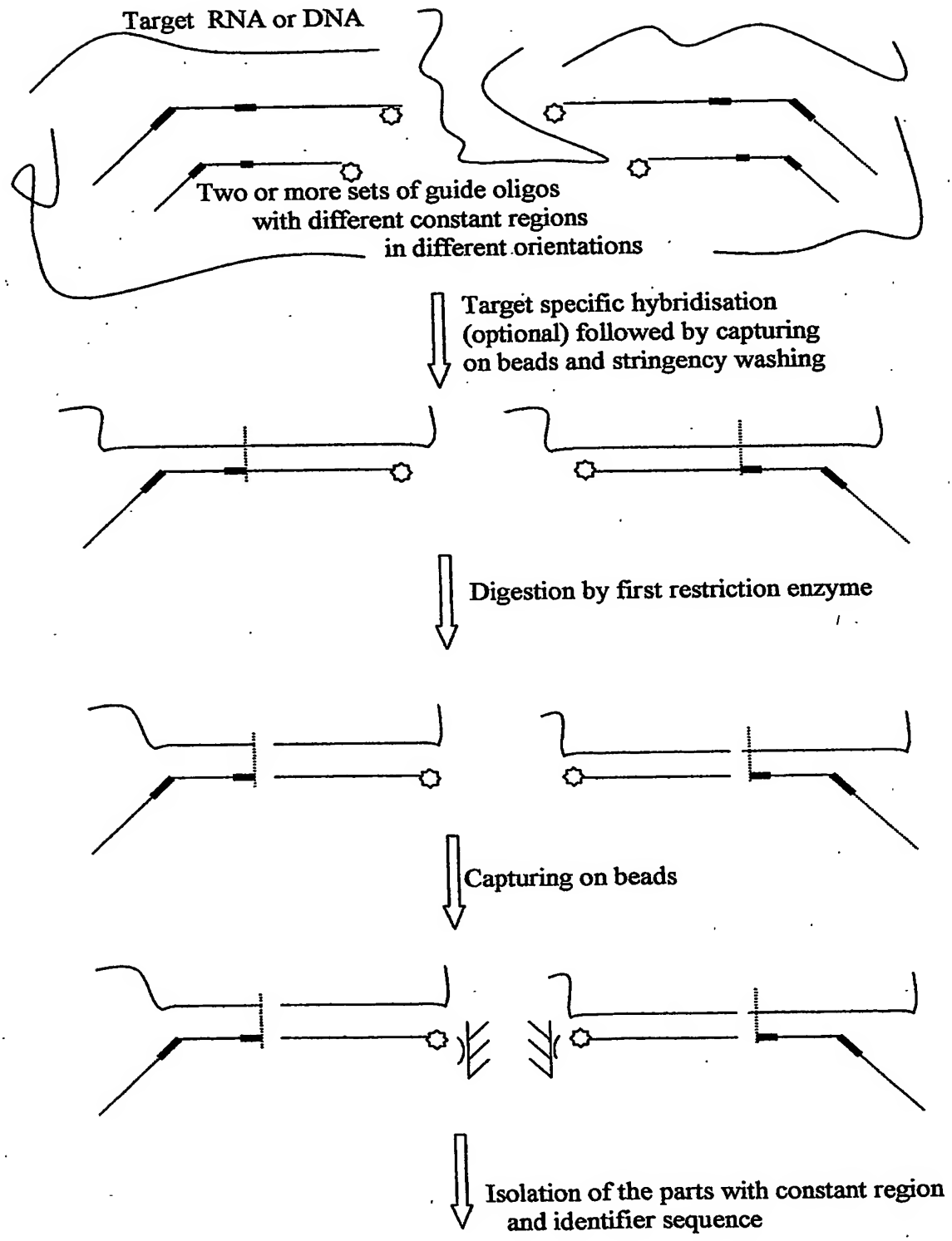


FIG. 3 (continuing)

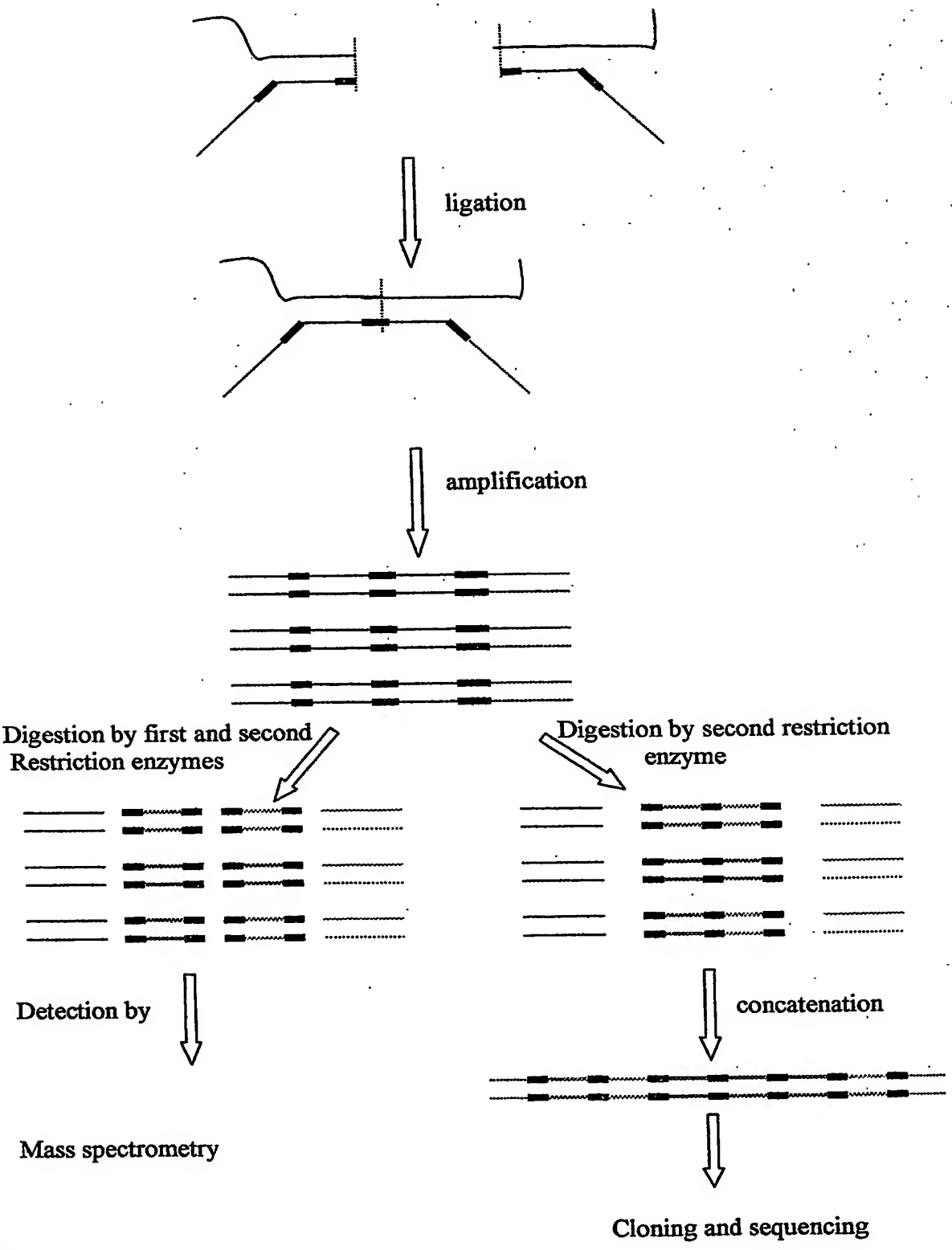

ligation
amplification
Digestion by first and second Restriction enzymes
Digestion by second restriction enzyme
Detection by
Mass spectrometry
concatenation
Cloning and sequencing

FIG. 3

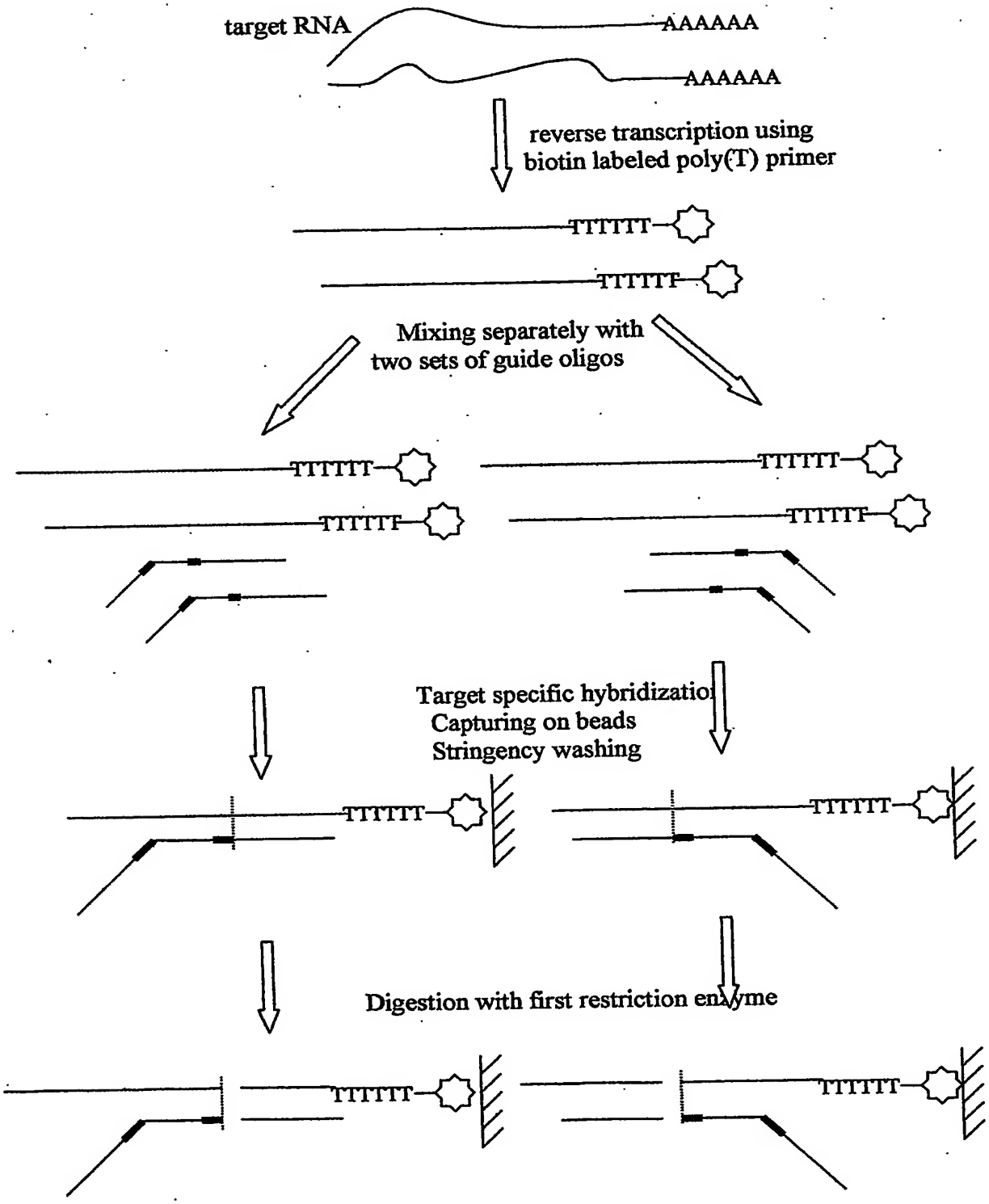
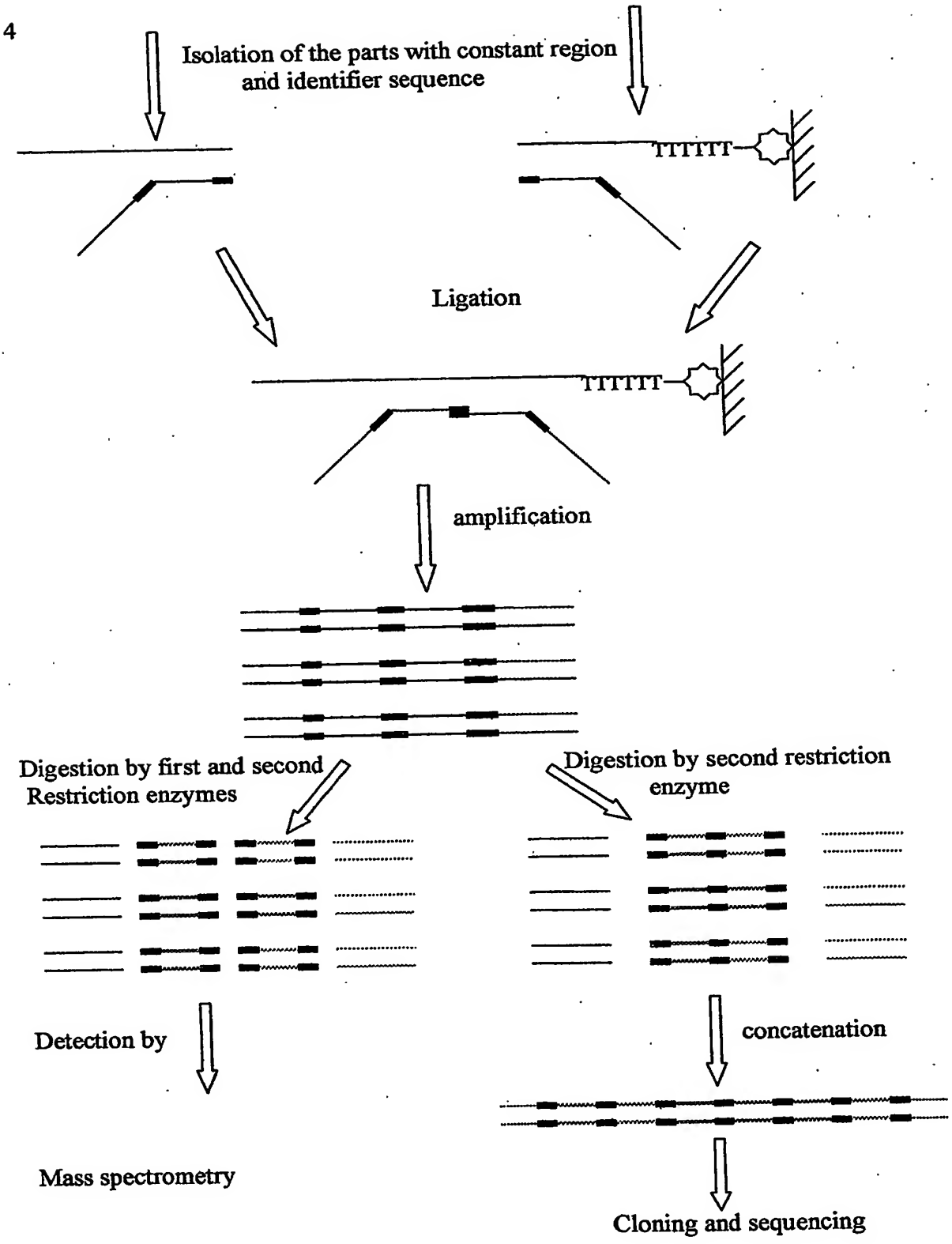


FIG. 4 (continuing)

FIG. 4


Isolation of the parts with constant region and identifier sequence
TTTTTT
Ligation
TTTTTT
amplification
Digestion by first and second Restriction enzymes
Digestion by second restriction enzyme
Detection by
Mass spectrometry
concatenation
Cloning and sequencing